

Faster Black-Box Algorithms Through Higher Arity Operators

Benjamin Doerr
Max-Planck-Institut für Informatik
Campus E1 4
66123 Saarbrücken, Germany

Timo Kötzing
Max-Planck-Institut für Informatik
Campus E1 4
66123 Saarbrücken, Germany*

Daniel Johannsen
Max-Planck-Institut für Informatik
Campus E1 4
66123 Saarbrücken, Germany

Per Kristian Lehre
DTU Informatics
Technical University of Denmark
2800 Lyngby, Denmark[†]

Markus Wagner
Max-Planck-Institut für Informatik
Campus E1 4
66123 Saarbrücken, Germany

Carola Winzen
Max-Planck-Institut für Informatik
Campus E1 4
66123 Saarbrücken, Germany[‡]

Abstract

We extend the work of Lehre and Witt (GECCO 2010) on the unbiased black-box model by considering higher arity variation operators.

*Timo Kötzing was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant NE 1182/5-1.

[†]Supported by Deutsche Forschungsgemeinschaft (DFG) under grant no. WI 3552/1-1.

[‡]Carola Winzen is a recipient of the Google Europe Fellowship in Randomized Algorithms, and this research is supported in part by this Google Fellowship.

In particular, we show that already for binary operators the black-box complexity of LEADINGONES drops from $\Theta(n^2)$ for unary operators to $O(n \log n)$. For ONEMAX, the $\Omega(n \log n)$ unary black-box complexity drops to $O(n)$ in the binary case. For k -ary operators, $k \leq n$, the ONEMAX-complexity further decreases to $O(n/\log k)$.

1 Introduction

When we analyze the optimization time of randomized search heuristics, we typically assume that the heuristic does not know anything about the objective function apart from its membership in a large class of functions, e.g., linear or monotone functions. Thus, the function is typically considered to be given as a black-box, i.e., in order to optimize the function, the algorithm needs to query the function values of various search points. The algorithm may then use the information on the function values to create new search points. We call the minimum number of function evaluations needed for a randomized search heuristic to optimize any function f of a given function class \mathcal{F} the *black-box complexity* of \mathcal{F} . We may restrict the algorithms with respect to how it creates new search points from the information collected in previous steps. Intuitively, the stronger restrictions that are imposed on which search points the algorithm can query next, the larger the black-box complexity of the function class.

Black-box complexity for search heuristics was introduced in 2006 by Droste, Jansen, and Wegener [DJW06]. We call their model the *unrestricted* black-box model as it imposes few restrictions on how the algorithm may create new search points from the information at hand. This model was the first attempt towards creating a complexity theory for randomized search heuristics. However, the authors prove bounds that deviate from those known for well-studied search heuristics, such as random local search and evolutionary algorithms. For example, the well-studied function class ONEMAX has an unrestricted black-box complexity of $\Theta(n/\log n)$ whereas standard search

Model	Arity	ONEMAX		LEADINGONES	
unbiased	1	$\Theta(n \log n)$	[LW10]	$\Theta(n^2)$	[LW10]
unbiased	$1 < k \leq n$	$O(n/\log k)$	(here)	$O(n \log n)$	(here)
unrestricted	n/a	$\Omega(n/\log n)$	[DJW06]	$\Omega(n)$	[DJW06]
		$O(n/\log n)$	[AW09]		

Table 1: Black-Box Complexity of ONEMAX and LEADINGONES. Note that upper bounds for the unbiased unary black-box complexity immediately carry over to higher arities. Similarly, lower bounds for the unrestricted black-box model also hold for the unbiased model.

heuristics only achieve a $\Omega(n \log n)$ runtime. Similarly, the class LEADING-ONES has a linear unrestricted black-box complexity but we typically observe a $\Omega(n^2)$ behavior for standard heuristics.

These gaps, among other reasons, motivated Lehre and Witt [LW10] to propose an alternative model. In their *unbiased* black-box model the algorithm may only invoke a so-called unbiased variation operator to create new search points. A variation operator returns a new search point given one or more previous search points. Now, intuitively, the unbiasedness condition implies that the variation operator is symmetric with respect to the bit values and bit positions. Or, to be more precise, it must be invariant under Hamming-automorphisms. We give a formal definition of the two black-box models in Section 2.

Among other problem instances, Lehre and Witt analyze the unbiased black-box complexity of the two function classes ONEMAX and LEADING-ONES. They can show that the complexity of ONEMAX and LEADING-ONES match the above mentioned $\Theta(n \log n)$ and, respectively, $\Theta(n^2)$ bounds, if we only allow *unary* operators. I.e., if the variation operator may only use the information from at most one previously queried search point, the unbiased black-box complexity matches the runtime of the well-known $(1 + 1)$ Evolutionary Algorithm.

In their first work, Lehre and Witt give no results on the black-box complexity of *higher arity* models. A variation operator is said to be of arity k if it creates new search points by recombining up to k previously queried search points. We are interested in higher arity black-box models because they include commonly used search heuristics which are not covered by the unary operators. Among such heuristics are evolutionary algorithms that employ uniform crossover, particle swarm optimization [KE01], ant colony optimization [DS04] and estimation of distribution algorithms [LL02].

Although search heuristics that employ higher arity operators are poorly understood from a theoretical point of view, there are some results proving that situations exist where higher arity is helpful. For example, Doerr, Klein, and Happ [DHK08] show that a concatenation operator reduces the runtime on the all-pairs shortest path problem. Refer to the same paper for further references.

Extending the work from Lehre and Witt, we analyze higher arity black-box complexities of ONEMAX and LEADING-ONES. In particular, we show that, surprisingly, the unbiased black-box complexity drops from $\Theta(n^2)$ in the unary case to $O(n \log n)$ for LEADING-ONES and from $\Theta(n \log n)$ to an at most linear complexity for ONEMAX. As the bounds for unbiased unary black-box complexities immediately carry over to all higher arity unbiased black-box complexities, we see that increasing the arity of the variation operators provably helps to decrease the complexity. We are optimistic that the ideas developed to prove the bounds can be further exploited to achieve reduced black-box complexities also for other function classes.

In this work, we also prove that increasing the arity further does again help. In particular, we show that for every $k \leq n$, the unbiased k -ary black-box complexity of ONEMAX can be bounded by $O(n/\log k)$. This bound is optimal for $k = n$, because the unbiased black-box complexity can always be bounded below by the unrestricted black-box complexity, which is known to be $\Omega(n/\log n)$ for ONEMAX [DJW06].

Note that a comparison between the unrestricted black-box complexity and the unbiased black-box complexity of LEADINGONES cannot be derived that easily. The asymptotic linear unrestricted black-box complexity mentioned above is only known to hold for a subclass of the class LEADINGONES considered in this work.

Table 1 summarizes the results obtained in this paper, and provides a comparison with known results on black-box complexity of ONEMAX and LEADINGONES.

2 Unrestricted and Unbiased Black-Box Complexities

In this section, we formally define the two black-box models by Droste, Jansen, and Wegener [DJW06], and Lehre and Witt [LW10]. We call the first model the *unrestricted black-box model*, and the second model the *unbiased black-box model*. Each model specifies a class of algorithms. The black-box complexity of a function class is then defined with respect to the algorithms specified by the corresponding model. We start by describing the two models, then provide the corresponding definitions of black-box complexity.

In both models, one is faced with a class of pseudo-Boolean functions \mathcal{F} that is known to the algorithm. An adversary chooses a function f from this class. The function f itself remains unknown to the algorithm. The algorithm can only gain knowledge about the function f by querying an oracle for the function value of search points. The goal of the algorithm is to find a globally optimal search point for the function. Without loss of generality, we consider maximization as objective. The two models differ in the information available to the algorithm, and the search points that can be queried.

Let us begin with some notation. Throughout this paper, we consider the maximization of pseudo-Boolean functions $f : \{0, 1\}^n \rightarrow \mathbb{R}$. In particular, n will always denote the length of the bitstring to be optimized. For a bitstring $x \in \{0, 1\}^n$, we write $x = x_1 \cdots x_n$. For convenience, we denote the positive integers by \mathbb{N} . For $k \in \mathbb{N}$, we use the notion $[k]$ as a shorthand for the set $\{1, \dots, k\}$. Analogously, we define $[0..k] := [k] \cup \{0\}$. Furthermore, let S_k denote the set of all permutations of $[k]$. With slight abuse of notation, we write $\sigma(x) := x_{\sigma(1)} \cdots x_{\sigma(n)}$ for $\sigma \in S_n$. Furthermore, the bitwise exclusive-

or is denoted by \oplus . For any bitstring x we denote its complement by \bar{x} . Finally, we use standard notation for asymptotic growth of functions (see, e.g., [CLRS01]). In particular, we denote by $o_n(g(n))$ the set of all functions f that satisfy $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$.

The *unrestricted black-box model* contains all algorithms which can be formalized as in Algorithm 1. A basic feature is that this scheme does not force any relationship between the search points of subsequent queries. Thus, this model contains a broad class of algorithms.

Algorithm 1: Unrestricted Black-Box Algorithm

- 1 Choose a probability distribution p_0 on $\{0, 1\}^n$.
 - 2 Sample x^0 according to p_0 and query $f(x^0)$.
 - 3 **for** $t = 1, 2, 3, \dots$ *until termination condition met* **do**
 - 4 Depending on $((x^0, f(x^0)), \dots, (x^{t-1}, f(x^{t-1})))$, choose
 - 5 a probability distribution p^t on $\{0, 1\}^n$.
 - 6 Sample x^t according to p^t , and query $f(x^t)$.
-

To exclude some algorithms whose behavior does not resemble those of typical search-heuristics, one can impose further restrictions. The *unbiased black-box model* introduced in [LW10] restricts Algorithm 1 in two ways. First, the decisions made by the algorithm only depends on the observed fitness values, and not the actual search points. Second, the algorithm can only query search points obtained by variation operators that are unbiased in the following sense. By imposing these two restrictions, the black-box complexity matches the runtime of popular search heuristics on example functions.

Definition 1. (UNBIASED k -ARY VARIATION OPERATOR [LW10]) *Let $k \in \mathbb{N}$. An unbiased k -ary distribution $D(\cdot \mid x^1, \dots, x^k)$ is a conditional probability distribution over $\{0, 1\}^n$, such that for all bitstrings $y, z \in \{0, 1\}^n$, and each permutation $\sigma \in S_n$, the following two conditions hold.*

- (i) $D(y \mid x^1, \dots, x^k) = D(y \oplus z \mid x^1 \oplus z, \dots, x^k \oplus z),$
- (ii) $D(y \mid x^1, \dots, x^k) = D(\sigma(y) \mid \sigma(x^1), \dots, \sigma(x^k)).$

An unbiased k -ary variation operator p is a k -ary operator which samples its output according to an unbiased k -ary distribution.

The first condition in Definition 1 is referred to as \oplus -invariance, and the second condition is referred to as *permutation invariance*. Note that the combination of these two conditions can be characterized as invariance under Hamming-automorphisms: $D(\cdot \mid x^1, \dots, x^k)$ is unbiased if and only if, for all $\alpha : \{0, 1\}^n \rightarrow \{0, 1\}^n$ preserving the Hamming distance and all

bitstrings y , $D(y \mid x^1, \dots, x^k) = D(\alpha(y) \mid \alpha(x^1), \dots, \alpha(x^k))$. We refer to 1-ary and 2-ary variation operators as *unary* and *binary* variation operators, respectively. The unbiased k -ary black-box model contains all algorithms which follow the scheme of Algorithm 2. While being a restriction of the old model, the unbiased model still captures the most widely studied search heuristics, including most evolutionary algorithms, simulated annealing and random local search.

Note that in line 5 of Algorithm 2, y^1, \dots, y^k don't necessarily have to be the k *immediately previously* queried ones. That is, the algorithm is allowed to choose any k previously sampled search points.

We now define black-box complexity formally. We will use query complexity as the cost model, where the algorithm is only charged for queries to the oracle, and all other computation is free. The runtime $T_{A,f}$ of a randomized algorithm A on a function $f \in \mathcal{F}$ is hence the expected number of oracle queries until the optimal search point is queried for the first time. The expectation is taken with respect to the random choices made by the algorithm.

Definition 2 (Black-box complexity). *The complexity of a class of pseudo-Boolean functions \mathcal{F} with respect to a class of algorithms \mathcal{A} , is defined as $T_{\mathcal{A}}(\mathcal{F}) := \min_{A \in \mathcal{A}} \max_{f \in \mathcal{F}} T_{A,f}$.*

The *unrestricted black-box complexity* is the complexity with respect to the algorithms covered by Algorithm 1. For any given $k \in \mathbb{N}$, the *unbiased k -ary black-box complexity* is the complexity with respect to the algorithms covered by Algorithm 2. Furthermore, the *unbiased $*$ -ary black-box complexity* is the complexity with respect to the algorithms covered by Algorithm 2, without limitation on the arity of the operators used.

It is easy to see that every unbiased k -ary operator p can be simulated by an unbiased $(k+1)$ -ary operator p' defined as $p'(z \mid x^1, \dots, x^k, x^{k+1}) := p(z \mid x^1, \dots, x^k)$. Hence, the unbiased k -ary black-box complexity is an upper bound for the unbiased $(k+1)$ -ary black-box complexity. Similarly, the set of unbiased black-box algorithms for any arity is contained in the set of unrestricted black-box algorithms. Therefore, the unrestricted black-box complexity is a lower bound for the unbiased k -ary black-box complexity (for all $k \in \mathbb{N}$).

3 The Unbiased $*$ -Ary Black-Box Complexity of OneMax

In this section, we show that the unbiased black-box complexity of ONEMAX is $\Theta(n/\log n)$ with a leading constant between one and two. We begin with the formal definition of the function class ONEMAX_n . We will omit the subscript “ n ” if the size of the input is clear from the context.

Algorithm 2: Unbiased k -ary Black-Box Algorithm

```

1 Sample  $x^0$  uniformly at random from  $\{0, 1\}^n$  and query  $f(x^0)$ .
2 for  $t = 1, 2, 3, \dots$  until termination condition met do
3   Depending on  $(f(x^0), \dots, f(x^{t-1}))$ , choose
4   an unbiased  $k$ -ary variation operator  $p^t$ , and
5    $k$  previously queried search points  $y^1, \dots, y^k$ .
6   Sample  $x^t$  according to  $p^t(y^1, \dots, y^k)$ , and query  $f(x^t)$ .
```

Definition 3 (ONEMAX). *For all $n \in \mathbb{N}$ and each $z \in \{0, 1\}^n$ we define $\text{OM}_z : \{0, 1\}^n \rightarrow \mathbb{N}, x \mapsto |\{j \in [n] \mid x_j = z_j\}|$.¹ The class ONEMAX_n is defined as $\text{ONEMAX}_n := \{\text{OM}_z \mid z \in \{0, 1\}^n\}$.*

To motivate the definitions, let us briefly mention that we do not further consider the optimization of specific functions such as $\text{OM}_{(1, \dots, 1)}$, since they would have an unrestricted black-box complexity of 1: The algorithm asking for the bitstring $(1, \dots, 1)$ in the first step easily optimizes the function in just one step. Thus, we need to consider some generalizations of these functions. For the unrestricted black-box model, we already have a lower bound by Droste Jansen, and Wegener [DJW06]. For the same model, an algorithm which matches this bound in order of magnitude is given by Anil and Wiegand in [AW09].

Theorem 4. *The unrestricted black-box complexity of ONEMAX_n is $\Theta(n/\log n)$. Moreover, the leading constant is at least 1.*

As already mentioned, the lower bound on the complexity of ONEMAX_n in the unrestricted black-box model from Theorem 4 directly carries over to the stricter unbiased black-box model.

Corollary 5. *The unbiased $*$ -ary black-box complexity of ONEMAX_n is at least $n/\log n$.*

Moreover, an upper bound on the complexity of ONEMAX in the unbiased black-box model can be derived using the same algorithmic approach as given for the unrestricted black-box model (compare [AW09] and Theorem 4).

Theorem 6. *The unbiased $*$ -ary black-box complexity of ONEMAX_n is at most $(1 + o_n(1)) \frac{2n}{\log n}$.*

In return, this theorem also applies to the unrestricted black-box model and refines Theorem 4 by explicitly bounding the leading constant of the *unrestricted* black-box complexity for ONEMAX by a factor of two of the

¹Intuitively, OM_z is the function of n minus the Hamming distance to z .

lower bound. The result in Theorem 6 is based on Algorithm 3. This algorithm makes use of the operator `uniformSample` that samples a bitstring uniformly at random, which clearly is an unbiased (0-ary) variation operator. Further, it makes use of another family of operators: `chooseConsistent` _{u^1, \dots, u^t} (x^1, \dots, x^t) chooses a $z \in \{0, 1\}^n$ uniformly at random such that, for all $i \in [t]$, $\text{OM}_z(x^i) = u^i$ (if there exists one, and any bitstring uniformly at random otherwise). It is easy to see that this is a family of unbiased variation operators.

Algorithm 3: Optimizing ONEMAX with unbiased variation operators.

```

1 input Integer  $n \in \mathbb{N}$  and function  $f \in \text{ONEMAX}_n$ ;
2 initialization  $t \leftarrow \lceil (1 + \frac{4 \log \log n}{\log n}) \frac{2n}{\log n} \rceil$ ;
3 repeat
4   foreach  $i \in [t]$  do
5      $x^i \leftarrow \text{uniformSample}()$ ;
6    $w \leftarrow \text{chooseConsistent}_{f(x^1), \dots, f(x^t)}(x^1, \dots, x^t)$ ;
7 until  $f(w) = n$ ;
8 output  $w$ ;
```

An upper bound of $(1 + o_n(1))2n/\log n$ for the expected runtime of Algorithm 3 follows directly from the following theorem which implies that the number of repetitions of steps 4 to 6 follows a geometric distribution with success probability $1 - o_n(1)$. This proves Theorem 6.

Theorem 7. *Let n be sufficiently large (i. e., let $n \geq N_0$ for some fixed constant $N_0 \in \mathbb{N}$). Let $z \in \{0, 1\}^n$ and let X be a set of $t \geq (1 + \frac{4 \log \log n}{\log n}) \frac{2n}{\log n}$ samples chosen from $\{0, 1\}^n$ uniformly at random and mutually independent. Then the probability that there exists an element $y \in \{0, 1\}^n$ such that $y \neq z$ and $\text{OM}_y(x) = \text{OM}_z(x)$ for all $x \in X$ is bounded from above by $2^{-t/2}$.*

The previous theorem is a refinement of Theorem 1 in [AW09], and its proof follows the proof of Theorem 1 in [AW09], clarifying some inconsistencies² in that proof. To show Theorem 7, we first give a bound on a combinatorial quantity used later in its proof (compare Lemma 1 in [AW09]).

Proposition 8. *For sufficiently large n ,*

$$t \geq \left(1 + \frac{4 \log \log n}{\log n}\right) \frac{2n}{\log n},$$

²For example, in the proof of Lemma 1 in [AW09] the following claim is made. Let $d(n)$ be a monotone increasing sequence that tends to infinity. Then for sufficient large n the sequence $h_d(n) = (\frac{\pi d(n)}{8})^{1/(2 \ln n)}$ is bounded away from 1 by a constant $b > 1$. Clearly, this is not the case. For example, for $d(n) = \log n$, the sequence $h_{\log}(n)$ converges to 1.

and even $d \in \{2, \dots, n\}$, it holds that

$$\binom{n}{d} \left(\binom{d}{d/2} 2^{-d} \right)^t \leq 2^{-3t/4}. \quad (1)$$

Proof. By Stirling's formula, we have $\binom{d}{d/2} \leq \left(\frac{\pi d}{2}\right)^{-1/2} 2^d$. Therefore,

$$\binom{n}{d} \left(\binom{d}{d/2} 2^{-d} \right)^t \leq \binom{n}{d} \left(\frac{\pi d}{2} \right)^{-t/2}. \quad (2)$$

We distinguish two cases. First, we consider the case $2 \leq d < n/(\log n)^3$. By Stirling's formula, it holds that $\binom{n}{d} \leq \left(\frac{en}{d}\right)^d$. Thus, we get from (2) that

$$\begin{aligned} \binom{n}{d} \left(\binom{d}{d/2} 2^{-d} \right)^t &\leq \left(\frac{en}{d}\right)^d \left(\frac{\pi d}{2}\right)^{-t/2} \\ &= 2^{\left(\frac{2d}{t} \log\left(\frac{en}{d}\right) - \log\left(\frac{\pi d}{2}\right)\right) \frac{t}{2}}. \end{aligned} \quad (3)$$

We bound d by its minimal value 2 and maximal value $n/(\log n)^3$, and t by $2n/\log n$ to obtain

$$\frac{2d}{t} \log \frac{en}{d} - \log \frac{\pi d}{2} \leq \frac{1}{(\log n)^2} \log \frac{en}{2} - \log \pi.$$

Since the first term on the right hand side converges to 0 and since $\log \pi > 3/2$, the exponent in (3) can be bounded from above by $-3t/4$, if n is sufficiently large. Thus, we obtain inequality (1) for $2 \leq d < n/(\log n)^3$.

Next, we consider the case $n/(\log n)^3 \leq d \leq n$. By the binomial formula, it holds that $\binom{n}{d} \leq 2^n$. Thus,

$$\binom{n}{d} \left(\frac{\pi d}{2} \right)^{-t/2} \leq 2^n \left(\frac{\pi d}{2} \right)^{-t/2} = 2^{\left(\frac{2n}{t} - \log \frac{\pi d}{2}\right) \frac{t}{2}}. \quad (4)$$

We bound $\pi d/2$ by $n/(\log n)^3$ and t by $(1 + \frac{4 \log \log n}{\log n}) \frac{2n}{\log n}$ to obtain

$$\begin{aligned} \frac{2n}{t} - \log \frac{\pi d}{2} &\leq \frac{\log n}{1 + \frac{4 \log \log n}{\log n}} - \log(n/(\log n)^3) \\ &= \frac{\log n}{1 + \frac{4 \log \log n}{\log n}} - \log n + 3 \log \log n \\ &= \frac{3 \log \log n + \frac{4 \log \log n}{\log n} (-\log n + 3 \log \log n)}{1 + \frac{4 \log \log n}{\log n}} \\ &= -\frac{\log n - 12 \log \log n}{\log n + 4 \log \log n} \log \log n. \end{aligned}$$

Again, for sufficiently large n the right hand side becomes smaller than $-3/2$. We combine the previous inequality with inequalities (2) and (4) to show inequality (1) for $n/(\log n)^3 \leq d \leq n$. \square

With the previous proposition at hand, we finally prove Theorem 7.

Proof of Theorem 7. Let n be sufficiently large, $z \in \{0, 1\}^n$, and X a set of $t \geq (1 + \frac{4 \log \log n}{\log n}) \frac{2n}{\log n}$ samples chosen from $\{0, 1\}^n$ uniformly at random and mutually independent.

For $d \in [n]$, let $A_d := \{y \in \{0, 1\}^n \mid n - \text{OM}_z(y) = d\}$ be the set of all points with Hamming distance d from z . Let $d \in [n]$ and $y \in A_d$. We say the point y is *consistent* with x if $\text{OM}_y(x) = \text{OM}_z(x)$ holds. Intuitively, this means that OM_y is a possible target function, given the fitness of x . It is easy to see that y is consistent with x if and only if x and y (and therefore x and z) differ in exactly half of the d bits that differ between y and z . Therefore, y is never consistent with x if d is odd and the probability that y is consistent with x is $\binom{d}{d/2} 2^{-d}$ if d is even.

Let p be the probability that there exists a point $y \in \{0, 1\}^n \setminus \{z\}$ such that y is consistent with all $x \in X$. Then,

$$p = \Pr \left(\bigcup_{y \in \{0, 1\}^n \setminus \{z\}} \bigcap_{x \in X} \text{"}y \text{ is consistent with } x\text{"} \right).$$

Thus, by the union bound, we have

$$p \leq \sum_{y \in \{0, 1\}^n \setminus \{z\}} \Pr \left(\bigcap_{x \in X} \text{"}y \text{ is consistent with } x\text{"} \right).$$

Since, for a fixed y , the events “ y is consistent with x ” are mutually independent for all $x \in X$, it holds that

$$p \leq \sum_{d=1}^n \sum_{y \in A_d} \prod_{x \in X} \Pr(\text{"}y \text{ is consistent with } x\text{"}).$$

We substitute the probability that a fixed $y \in \{0, 1\}^n$ is consistent with a randomly chosen $x \in \{0, 1\}^n$ as given above. Using $|A_d| = \binom{n}{d}$, we obtain

$$p \leq \sum_{d \in \{1, \dots, n\}: d \text{ even}} \binom{n}{d} \left(\binom{d}{d/2} 2^{-d} \right)^t$$

Finally, we apply Proposition 8 and have $p \leq n 2^{-3t/4}$ which concludes the proof since $n \leq 2^{t/4}$ for sufficiently large n (as t in $\Omega(n/\log n)$). \square

4 The Unbiased k -Ary Black-Box Complexity of OneMax

In this section, we show that higher arity indeed enables the construction of faster black-box algorithms. In particular, we show the following result.

Theorem 9. *For every $k \in [n]$ with $k \geq 2$, the unbiased k -ary black-box complexity of ONEMAX_n is at most linear in n . Moreover, it is at most $(1 + o_k(1))2n/\log k$.*

This result is surprising, since in [LW10], Lehre and Witt prove that the unbiased unary black-box complexity of the class of all functions f with a unique global optimum is $\Omega(n \log n)$. Thus, we gain a factor of $\log n$ when switching from unary to binary variation operators.

To prove Theorem 9, we introduce two different algorithms interesting on their own. Both algorithms share the idea to track which bits have already been optimized. That way we can avoid flipping them again in future iterations of the algorithm.

The first algorithm proves that the unbiased binary black-box complexity of ONEMAX_n is at most linear in n if the arity is at least two. For the general case, with $k \geq 3$, we give a different algorithm that provides asymptotically better bounds for k growing in n . We use the idea that the whole bitstring can be divided into smaller substrings, and subsequently those can be independently optimized. We show that this is possible, and together with Theorem 6, this yields the above bound for ONEMAX_n in the k -ary case for $k \geq 3$.

4.1 The Binary Case

We begin with the binary case. We use the three unbiased variation operators `uniformSample` (as described in Section 3), `complement` and `flipOneWhereDifferent` defined as follows. The unary operator `complement(x)` returns the bitwise complement of x . The binary operator `flipOneWhereDifferent(x, y)` returns a copy of x , where one of the bits that differ in x and y is chosen uniformly at random and then flipped. It is easy to see that `complement` and `flipOneWhereDifferent` are unbiased variation operators.

Lemma 10. *With exponentially small probability of failure, the optimization time of Algorithm 4 on the class ONEMAX_n is at most $(1+\varepsilon)2n$, for all $\varepsilon > 0$. The algorithm only involves binary operators.*

Proof. We first prove that the algorithm is correct. Assume that the instance has optimum z , for some $z \in \{0, 1\}^n$. We show that the following invariant is satisfied in the beginning of every iteration of the main loop (steps 4-12): for all $i \in [n]$, if $x_i = y_i$, then $x_i = z_i$. In other words, the positions where x and y have the same bit value are optimized. The invariant clearly holds in the first iteration, as x and y differ in all bit positions. A bit flip is only accepted if the fitness value is strictly higher, an event which occurs with positive probability. Hence, if the invariant holds in the current iteration, then it also holds in the following iteration. By induction, the invariant property now holds in every iteration of the main loop.

Algorithm 4: Optimizing ONEMAX with unbiased binary variation operators.

```

1 input Integer  $n \in \mathbb{N}$  and function  $f \in \text{ONEMAX}_n$ ;
2 initialization  $x \leftarrow \text{uniformSample}()$ ;
3  $y \leftarrow \text{complement}(x)$ ;
4 repeat
5   Choose  $b \in \{0, 1\}$  uniformly at random;
6   if  $b = 1$  then
7      $x' \leftarrow \text{flipOneWhereDifferent}(x, y)$ ;
8     if  $f(x') > f(x)$  then  $x \leftarrow x'$ ;
9   else
10     $y' \leftarrow \text{flipOneWhereDifferent}(y, x)$ ;
11    if  $f(y') > f(y)$  then  $y \leftarrow y'$ ;
12 until  $f(x) = n$ ;
13 output  $x$ ;

```

We then analyze the runtime of the algorithm. Let T be the number of iterations needed until n bit positions have been optimized. Due to the invariant property, this is the same as the time needed to reduce the Hamming distance between x and y from n to 0. An iteration is successful, i.e., the Hamming distance is reduced by 1, with probability $1/2$ independently of previous trials. The random variable T is therefore negative binomially distributed with parameters n and $1/2$. It can be related to a binomially distributed random variable X with parameters $2n(1 + \varepsilon)$ and $1/2$ by $\Pr(T \geq 2n(1 + \varepsilon)) = \Pr(X \leq n)$. Finally, by applying a Chernoff bound with respect to X , we obtain $\Pr(T \geq 2n(1 + \varepsilon)) \leq \exp(-\varepsilon^2 n / 2(1 + \varepsilon))$. \square

It is easy to see that Algorithm 4 yields the same bounds on the class of monotone functions, which is defined as follows.

Definition 11 (Monotone functions). *Let $n \in \mathbb{N}$ and let $z \in \{0, 1\}^n$. A function $f : \{0, 1\}^n \rightarrow \mathbb{R}$ is said to be monotone with respect to z if for all $y, y' \in \{0, 1\}^n$ with $\{i \in [n] \mid y_i = z_i\} \subsetneq \{i \in [n] \mid y'_i = z_i\}$ it holds that $f(y) < f(y')$. The class MONOTONE_n contains all such functions that are monotone with respect to some $z \in \{0, 1\}^n$.*

Now, let f be a monotone function with respect to z and let y and y' be two bitstrings which differ only in the i -th position. Assume that $y_i \neq z_i$ and $y'_i = z_i$. It follows from the monotonicity of f that $f(y) < f(y')$. Consequently, Algorithm 4 optimizes f as fast as any function in ONEMAX_n .

Corollary 12. *The unbiased binary black-box complexity of MONOTONE_n is $O(n)$.*

Note that MONOTONE_n strictly includes the class of linear functions with non-zero weights.

4.2 Proof of Theorem 9 for Arity $k \geq 3$

For the case of arity $k \geq 3$ we analyze the following Algorithm 5 and show that its optimization time on ONEMAX_n is at most $(1 + o_k(1))2n/\log k$. Informally, the algorithm splits the bitstring into blocks of length k . The n/k blocks are then optimized separately using a variant of Algorithm 3, each in expected time $(1 - o_k(1))2k/\log k$.

In detail, Algorithm 5 maintains its state using three bitstrings x, y and z . Bitstring x represents the preliminary solution. The positions in which bitstrings x and y differ represent the remaining blocks to be optimized, and the positions in which bitstrings y and z differ represent the current block to be optimized. Due to permutation invariance, it can be assumed without loss of generality that the bitstrings can be expressed by $x = \alpha\bar{\beta}\gamma$, $y = \bar{\alpha}\beta\gamma$, and $z = \alpha\beta\gamma$, see Step 6 of Algorithm 5. The algorithm uses an operator called $\text{flipKWhereDifferent}_\ell$ to select a new block of size ℓ to optimize. The selected block is optimized by calling the subroutine $\text{optimizeSelected}_{n,\ell}$, and the optimized block is inserted into the preliminary solution using the operator update .

The operators in Algorithm 5 are defined as follows. The operator $\text{flipKWhereDifferent}_k(x, y)$ generates the bitstring z . This is done by making a copy of y , choosing $\ell := \min\{k, H(x, y)\}$ bit positions for which x and y differ uniformly at random, and flipping them. The operator $\text{update}(a, b, c)$ returns a bitstring a' which in each position $i \in [n]$ independently, takes the value $a'_i = b_i$ if $a_i = c_i$, and $a'_i = a_i$ otherwise. Clearly, both these operators are unbiased. The operators uniformSample and complement have been defined in previous sections.

It remains to define the subroutine $\text{optimizeSelected}_{n,k}$. This subroutine is a variant of Algorithm 3 that only optimizes a selected block of bit positions, and leaves the other blocks unchanged. The block is represented by the bit positions in which bitstrings y and z differ. Due to permutation-invariance, we assume that they are of the form $y = \bar{\alpha}\sigma$ and $z = \alpha\sigma$, for some bitstrings $\alpha \in \{0, 1\}^k$, and $\sigma \in \{0, 1\}^{n-k}$. The operator uniformSample in Algorithm 3 is replaced by a 2-ary operator defined by: $\text{randomWhereDifferent}(x, y)$ chooses z , where for each $i \in [n]$, the value of bit z_i is x_i or y_i with equal probability. Note that this operator is the same as the standard uniform crossover operator. The operator family chooseConsistent in Algorithm 3 is replaced by a family of $(r + 2)$ -ary operators defined by: $\text{chooseConsistentSelected}_{u^1, \dots, u^r}(x^1, \dots, x^r, \bar{\alpha}\sigma, \alpha\sigma)$ chooses $z\sigma$, where the prefix z is sampled uniformly at random from the set $Z_{u,x} = \{z \in \{0, 1\}^k \mid \forall i \in [t] \text{ OM}_z(x_1^i x_2^i \dots x_k^i) = u^i\}$. If the set $Z_{u,x}$ is empty, then z is sampled uniformly at random among all bitstrings of

Algorithm 5: Optimizing ONEMAX with unbiased k -ary variation operators, for $k \geq 3$.

```

1 input Integers  $n, k \in \mathbb{N}$ , and function  $f \in \text{ONEMAX}_n$ ;
2 initialization  $x^1 \leftarrow \text{uniformSample}()$ ,  $y^1 \leftarrow \text{complement}(x)$ , and
    $\tau \leftarrow \lceil \frac{n}{k} \rceil$ ;
3 foreach  $t \in [\tau]$  do
4    $\ell(t) \leftarrow \min\{k, n - k(t-1)\}$ ;
5    $z \leftarrow \text{flipKWhereDifferent}_{\ell(t)}(x^t, y^t)$ ;
6   Assume that  $x^t = \alpha\bar{\beta}\gamma$ ,  $y^t = \bar{\alpha}\beta\gamma$ , and  $z = \alpha\beta\gamma$ ;
7    $w^t\beta\gamma \leftarrow \text{optimizeSelected}_{n, \ell(t)}(\bar{\alpha}\beta\gamma, \alpha\beta\gamma)$ ;
8    $w^t\bar{\beta}\gamma \leftarrow \text{update}(\alpha\bar{\beta}\gamma, w^t\beta\gamma, \alpha\beta\gamma)$ ;
9    $x^{t+1} \leftarrow w^t\bar{\beta}\gamma$  and  $y^{t+1} \leftarrow w^t\beta\gamma$ ;
10 output  $x^{\tau+1}$ ;
```

length k . Informally, the set $Z_{u,x}$ corresponds to the subset of functions in ONEMAX_n that are consistent with the function values u^1, u^2, \dots, u^r on the inputs x^1, x^2, \dots, x^r . It is easy to see that this operator is unbiased.

Algorithm 6: `optimizeSelected` used in Algorithm 5.

```

1 input Integers  $n, k \in \mathbb{N}$ , and bitstrings  $\bar{\alpha}\sigma$  and  $\alpha\sigma$ , where  $\alpha \in \{0, 1\}^k$ 
   and  $\sigma \in \{0, 1\}^{n-k}$ ;
2 initialization  $r \leftarrow \min\left\{k-2, \left\lceil \left(1 + \frac{4 \log \log k}{\log k}\right) \frac{2k}{\log k} \right\rceil\right\}$ ,
    $f_\sigma \leftarrow \frac{f(\alpha\sigma) + f(\bar{\alpha}\sigma) - k}{2}$ ;
3 repeat
4   foreach  $i \in [r]$  do
5      $x^i\sigma \leftarrow \text{randomWhereDifferent}(\alpha\sigma, \bar{\alpha}\sigma)$ ;
6      $w\sigma \leftarrow$ 
        $\text{chooseConsistentSelected}_{f(x^1\sigma)-f_\sigma, \dots, f(x^r\sigma)-f_\sigma}(x^1\sigma, \dots, x^r\sigma, \bar{\alpha}\sigma, \alpha\sigma)$ ;
7 until  $f(w\sigma) = k + f_\sigma$ ;
8 output  $w\sigma$ ;
```

Proof of Theorem 9 for arity $k \geq 3$. To prove the *correctness* of the algorithm, assume without loss of generality the input $f = \text{ONEMAX}$ for which the correct output is 1^n .

We first claim that a call to $\text{optimizeSelected}_{n,k}(\bar{\alpha}\sigma, \alpha\sigma)$ will terminate after a finite number of iterations with output $1^k\sigma$ almost surely. The variable f_σ is assigned in line 2 of Algorithm 6, and it is easy to see that it takes the value $f_\sigma = f(0^k\sigma)$. It follows from linearity of f

and from $f(\alpha 0^{n-k}) + f(\bar{\alpha} 0^{n-k}) = k$, that $f(w 0^{n-k}) = f(w\sigma) - f(0^k\sigma) = f(w\sigma) - f_\sigma$. The termination condition $f(w\sigma) = k + f_\sigma$ is therefore equivalent to the condition $w\sigma = 1^k\sigma$. For all $x \in \{0,1\}^k$, it holds that $\text{OM}_{(1,\dots,1)}(x) = f(x)$, so 1^k is member of the set $Z_{u,x}$. Hence, every invocation of `chooseConsistentSelected` returns $1^k\sigma$ with non-zero probability. Therefore, the algorithm terminates after every iteration with non-zero probability, and the claim holds.

We then prove by induction the invariant property that for all $t \in [\tau+1]$, and $i \in [n]$, if $x_i^t = y_i^t$ then $x_i^t = y_i^t = 1$. The invariant clearly holds for $t = 1$, so assume that the invariant also holds for $t = j \leq \tau$. Without loss of generality, $x^j = \alpha\bar{\beta}\gamma$, $y^j = \bar{\alpha}\beta\gamma$, $x^{j+1} = w^j\bar{\beta}\gamma$, and $y^{j+1} = w^j\beta\gamma$. By the claim above and the induction hypothesis, both the common prefix w^j and the common suffix γ consist of only 1-bits. So the invariant holds for $t = j + 1$, and by induction also for all $t \in [\tau + 1]$.

It is easy to see that for all $t \leq \tau$, the Hamming distance between $x^{t+1} = w^t\bar{\beta}\gamma$ and $y^{t+1} = w^t\beta\gamma$ is $H(x^{t+1}, y^{t+1}) = H(\alpha\bar{\beta}\gamma, \bar{\alpha}\beta\gamma) - \ell(t) = H(x^t, y^t) - \ell(t)$. By induction, it therefore holds that

$$\begin{aligned} H(x^{\tau+1}, y^{\tau+1}) &= H(x^1, y^1) - \sum_{t=1}^{\tau} \ell(t) \\ &= n - \sum_{t=1}^{\tau} \min\{k, n - k(t-1)\} = 0. \end{aligned}$$

Hence, by the invariant above, the algorithm returns the correct output $x^{\tau+1} = y^{\tau+1} = 1^n$.

The *runtime* of the algorithm in each iteration is dominated by the subroutine `optimizeSelected`. Note that by definition, the probability that `chooseConsistentSelected` _{$f(x^1\sigma)-f_\sigma, \dots, f(x^r\sigma)-f_\sigma$} $(x^1\sigma, \dots, x^r\sigma, \bar{\alpha}\sigma, \alpha\sigma)$ chooses $z\sigma$ in $\{0,1\}^n$, is the same as the probability that `chooseConsistent` _{$f(x^1), \dots, f(x^r)$} (x^1, \dots, x^r) chooses z in $\{0,1\}^k$. To finish the proof, we distinguish between two cases.

Case 1: $k \leq 53$. In this case, it suffices³ to prove that the runtime is $O(n)$. For the case $k = 2$, this follows from Lemma 10. For $2 < k \leq 53$, it holds that $r = k - 2$. Each iteration in `optimizeSelected` uses $r + 1 = k - 1 = O(1)$ function evaluations, and the probability that `chooseConsistentSelected` optimizes a block of k bits is at least $1 - (1 - 2^{-k})^r = \Omega(1)$ (when $w = 1^k$). Thus, the expected optimization time for a block is $O(1)$, and for the entire bitstring it is at most $(n/k) \cdot O(1)$.

Case 2: $k \geq 54$. In this case, $r = \lceil (1 + \frac{4 \log \log k}{\log k}) \frac{2k}{\log k} \rceil$ holds. Hence, with an analysis analogous to that in the proof of Theorem 6, we can show

³ Assume that the expected runtime is less than cn for some constant $c > 0$ when $k \leq 53$. It is necessary to show that $cn \leq 2n/\log k + h(k)2n/\log k$, for some function h , where $\lim_{k \rightarrow \infty} h(k) \rightarrow 0$. This can easily be shown by choosing any such function h , where $h(k) \geq c \log k/2$ for $k \leq 53$.

that the expected runtime of `optimizeSelected` is at most $(1 + o_k(1))2(k - 2)/\log(k - 2)$. Thus, the expected runtime is at most $n/k \cdot (1 + o_k(1))2(k - 2)/\log(k - 2) = (1 + o_k(1))2n/\log k$. \square

5 The Complexity of LeadingOnes

In this section, we show that allowing k -ary variation operators, for $k > 1$, greatly reduces the black-box complexity of the `LEADINGONES` functions class, namely from $\Theta(n^2)$ down to $O(n \log n)$. We define the class `LEADINGONES` as follows.

Definition 13 (`LEADINGONES`). *Let $n \in \mathbb{N}$. Let $\sigma \in S_n$ be a permutation of the set $[n]$ and let $z \in \{0, 1\}^n$. The function $\text{LO}_{z,\sigma}$ is defined via $\text{LO}_{z,\sigma}(x) := \max\{i \in [0..n] \mid z_{\sigma(i)} = x_{\sigma(i)}\}$. We set $\text{LEADINGONES}_n := \{\text{LO}_{z,\sigma} \mid z \in \{0, 1\}^n, \sigma \in S_n\}$.*

The class `LEADINGONES` is well-studied. Already in 2002, Droste, Jansen and Wegener [DJW02] proved that the classical $(1 + 1)$ EA has an expected optimization time of $\Theta(n^2)$ on `LEADINGONES`. This bound seems to be optimal among the commonly studied versions of evolutionary algorithms. In [LW10], the authors prove that the unbiased unary black-box complexity of `LEADINGONES` is $\Theta(n^2)$.

Droste, Jansen and Wegener [DJW06] consider a subclass of LEADINGONES_n , namely $\text{LEADINGONES}_n^0 := \{\text{LO}_{z,\text{id}} \mid z \in \{0, 1\}^n\}$, where id denotes the identity mapping on $[n]$. Hence their function class is not permutation invariant. In this restricted setting, they prove a black-box complexity of $\Theta(n)$. Of course, their lower bound of $\Omega(n)$ is a lower bound for the unrestricted black-box complexity of the general LEADINGONES_n class, and consequently, a lower bound also for the unbiased black-box complexities of this class.

The following theorem is the main result in this section.

Theorem 14. *The unbiased binary black-box complexity of LEADINGONES_n is $O(n \log n)$.*

The key ingredient of the two black-box algorithms that yield our upper bound is an emulation of a binary search which determines the (unique) bit that increases the fitness *and* does flip this bit. Surprisingly, this can be done already with a binary operator. This works in spite of the fact that we also follow the general approach of the previous section of keeping two individuals x and y such that for all bit positions in which x and y agree, the corresponding bit value equals the one of the optimal solution.

We will use the two unbiased binary variation operators `randomWhereDifferent` (as described in Section 4.2) and

switchIfDistanceOne. The operator `switchIfDistanceOne(y, y')` returns y' if y and y' differ in exactly one bit, and returns y otherwise. It is easy to see that `switchIfDistanceOne` is an unbiased variation operators.

We call a pair (x, y) of search points *critical*, if the following two conditions are satisfied. (i) $f(x) \geq f(y)$. (ii) There are exactly $f(y)$ bit-positions $i \in [n]$ such that $x_i = y_i$. The following is a simple observation.

Lemma 15. *Let $f \in \text{LEADINGONES}_n$. If (x, y) is a critical pair, then either $f(x) = n = f(y)$ or $f(x) > f(y)$.*

If $f(x) > f(y)$, then the unique bit-position k such that flipping the k -th bit in x reduces its fitness to $f(y)$ – or equivalently, the unique bit-position such that flipping this bit in y increases y 's fitness – shall be called the *critical bit-position*. We also call $f(y)$ the value of the pair (x, y) .

Note that the above definition does only use some function values of f , but not the particular definition of f . If $f = \text{LO}_{\sigma, z}$, then the above implies that x and y are equal on the bit-positions $\sigma(1), \dots, \sigma(f(y))$ and are different on all other bit-positions. Also, the critical bit-position is $\sigma(f(y) + 1)$, and the only way to improve the fitness of y is flipping this particular bit-position (and keeping the positions $\sigma(1), \dots, \sigma(f(y))$ unchanged). The central part of Algorithm 7, which is contained in lines 3 to 9, manages to transform a critical pair of value $v < n$ into one of value $v + 1$ in $O(\log n)$ time. This is analyzed in the following lemma.

Lemma 16. *Assume that the execution of Algorithm 7 is before line 4, and that the current value of (x, y) is a critical pair of value $v < n$. Then after an expected number of $O(\log n)$ iterations, the loop in lines 5-9 is left and (x, y) or (y, x) is a critical pair of value $v + 1$.*

Proof. Let k be the critical bit-position of the pair (x, y) . Let $y' = x$ be a copy of x . Let $J := \{i \in [n] \mid y_i \neq y'_i\}$. Our aim is to flip all bits of y' with index in $J \setminus \{k\}$.

We define y'' by flipping each bit of y' with index in J with probability $1/2$. Equivalently, we can say that y''_i equals y'_i for all i such that $y'_i = y_i$, and is random for all other i (thus, we obtain such y'' by applying `randomWhereDifferent(y, y')`).

With probability exactly $1/2$, the critical bit was not flipped (“success”), and consequently, $f(y'') > f(y)$. In this case (due to independence), each other bit with index in J has a chance of $1/2$ of being flipped. So with constant probability at least $1/2$, $\{i \in [n] \mid y_i \neq y''_i\} \setminus \{k\}$ is at most half the size of $J \setminus \{k\}$. In this success case, we take y'' as new value for y' .

In consequence, the cardinality of $J \setminus \{k\}$ does never increase, and with probability at least $1/4$, it decreases by at least 50%. Consequently, after an expected number of $O(\log n)$ iterations, we have $|J| = 1$, namely $J = \{k\}$. We check this via an application of `switchIfDistanceOne`. \square

We are now ready to prove the main result of this section.

Proof of Theorem 14. We regard the following invariant: (x, y) or (y, x) is a critical pair. This is clearly satisfied after execution of line 1. From Lemma 16, we see that a single execution of the outer loop does not dissatisfy our invariant. Hence by Lemma 15, our algorithm is correct (provided it terminates). The algorithm does indeed terminate, namely in $O(n \log n)$ time, because, again by Lemma 16, each iteration of the outer loop increases the value of the critical pair by one. \square

Algorithm 7: Optimizing LEADINGONES with unbiased binary variation operators.

```

1 initialization  $x \leftarrow \text{uniformSample}(); y \leftarrow \text{complement}(x);$ 
2 repeat
3   if  $f(y) > f(x)$  then  $(x, y) \leftarrow (y, x);$ 
4    $y' \leftarrow x;$ 
5   repeat
6      $y'' \leftarrow \text{randomWhereDifferent}(y, y');$ 
7     if  $f(y'') > f(y)$  then  $y' \leftarrow y'';$ 
8      $y \leftarrow \text{switchIfDistanceOne}(y, y');$ 
9   until  $f(y) = f(y');$ 
10 until  $f(x) = f(y);$ 
11 output  $x;$ 

```

6 Conclusion and Future Work

We continue the study of the unbiased black-box model introduced in [LW10]. For the first time, we analyze variation operators with arity higher than one. Our results show that already two-ary operators can allow significantly faster algorithms.

The problem ONEMAX cannot be solved in shorter time than $\Omega(n \log n)$ with unary variation operators [LW10]. However, the runtime can be reduced to $O(n)$ with binary operators. The runtime can be decreased even further with higher arities than two. For k -ary variation operators, $2 \leq k \leq n$, the runtime can be reduced to $O(n/\log k)$, which for $k = n^{\Theta(1)}$ matches the lower bound in the classical black-box model. A similar positive effect of higher arity variation operators can be observed for the function class LEADINGONES. While this function class cannot be optimized faster than $\Omega(n^2)$ with unary variation operators [LW10], we show that the runtime can be reduced to $O(n \log n)$ with binary, or higher arity variation operators.

Despite the restrictions imposed by the unbiasedness conditions, our analysis demonstrates that black-box algorithms can employ new and more efficient search heuristics with higher arity variation operators. In particular, binary variation operators allow a memory mechanism that can be used to implement binary search on the positions in the bitstring. The algorithm can thereby focus on parts of the bitstring that has not previously been investigated.

An important open problem arising from this work is to provide lower bounds in the unbiased black-box model for higher arities than one. Due to the greatly enlarged computational power of black-box algorithms using higher arity operators (as seen in this paper), proving lower bounds in this model seems significantly harder than in the unary model.

References

- [AW09] G. Anil and R. P. Wiegand, *Black-box search by elimination of fitness functions*, Proc. of Foundations of Genetic Algorithms (FOGA'09), ACM, 2009, pp. 67–78.
- [CLRS01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed., McGraw Hill, 2001.
- [DHK08] B. Doerr, E. Happ, and C. Klein, *Crossover can provably be useful in evolutionary computation*, Proc. of Genetic and Evolutionary Computation Conference (GECCO '08), ACM, 2008, pp. 539–546.
- [DJW02] S. Droste, T. Jansen, and I. Wegener, *On the analysis of the $(1+1)$ evolutionary algorithm*, Theoretical Computer Science **276** (2002), 51–81.
- [DJW06] ———, *Upper and lower bounds for randomized search heuristics in black-box optimization*, Theoretical Computer Science **39** (2006), 525–544.
- [DS04] M. Dorigo and T. Stützle, *Ant colony optimization*, MIT Press, 2004.
- [KE01] J. Kennedy and R. C. Eberhart, *Swarm intelligence*, Morgan Kaufmann Publishers Inc., 2001.
- [LL02] P. Larrañaga and J. A. Lozano, *Estimation of distribution algorithms: a new tool for evolutionary computation*, Kluwer Academic Publishers, 2002.

- [LW10] P. K. Lehre and C. Witt, *Black-box search by unbiased variation*, Proc. of Genetic and Evolutionary Computation Conference (GECCO '10), ACM, 2010, pp. 1441–1448.